Improving ecological connectivity assessments with transfer learning and function approximation

Michael D. Catchen[1,2], **Michelle Lin**[1,3], **Timothée Poisot**[2,4], **David Rolnick**[1,3], **Andrew Gonzalez**[1,2]
[1]McGill University, [2]Quebec Centre for Biodiversity Science, [3]Mila — Quebec AI Institute
[4]Université de Montréal
{michael.catchen,andrew.gonzalez}@mcgill.ca
{michelle.lin, drolnick}@mila.quebec
timothee.poisot@umontreal.ca

Abstract

Protecting and restoring ecological connectivity is essential to climate change adaptation, and necessary if species are to shift their geographic distributions to track their suitable climatic conditions over the coming century. Despite the increasing availability of near real-time and high resolution data for landcover change, current connectivity planning projects are hindered by the computational time required to run connectivity analyses at realistic geographic scales with realistic models of movement. This bottleneck precludes application of optimization algorithms to prioritize ecological restoration to maintain and improve connectivity. Here we propose we can make progress toward overcoming these challenges using machine-learning methods. Our proposed methods will enable rapid optimization of connectivity prioritization and extend its application to many more species than is currently possible. We conclude by illustrating how this project will contribute to efforts to apply connectivity conservation using an example of ongoing restoration in southern Québec.

1 Problem and Motivation

The surface of Earth is changing (largely due to urban sprawl, agriculture, and deforestation), creating the process of *landscape fragmentation*, where landscapes become increasingly patchy. Fragmentation affects how organisms move through their habitat [1], and the ability for biomass to move across space is necessary for ecosystem functioning [2]. As a result, there is widespread interest in restoration efforts to promote *ecological connectivity* to maintain ecosystem function and to ensure organisms can disperse across landscapes to track their changing climatic niches [3]. Ensuring ecological connectivity is a primary goal of the UN Convention on Biological Diversity's Global Biodiversity Framework [4], and the 2030 target for this effort reflects the urgency to counter the impacts of human land use and climate change on biodiversity and ecosystems [5] by protecting and restoring connectivity.

Our ability to measure the structure and change of landscapes has rapidly improved in recent years due to the explosion in ecologically relevant satellite-based remote-sensing data [6], e.g. near real-time data of Earth's landcover [7]. However, our ability to quickly and reliably assess connectivity has not kept pace with this rapid increase in data availability. This landcover data can vastly improve our predictions of species movement, as the propensity for species to move between patches is not just a function of the distance between patches, but related to the entire context of the landscape around them [8]. The state-of-the-art model in this category is Circuitscape (CS) [9, 10], which produces pairwise estimates of movement flows between *patches* (core areas of habitat).

Problem Bottleneck

Despite being the best model of connectivity for many species, there are two major roadblocks that prevent Circuitscape from being used for rapid assessments of landscape connectivity to guide habitat restoration efforts. First, Circuitscape requires species-level resistance-to-movement (RTM) data, which is scarce. RTM refers to the relative obstruction a given landcover type poses to movement for a particular species, e.g. RTM for wolves is low in forests, moderate in agricultural land, and high in urban development (for a review of methods for collecting this data, see [11]). This lack of data inhibits our ability to quantify connectivity for most species, which is a significant gap in knowledge in landscape ecology [12]. Second, Circuitscape is too computationally intensive to be used in a framework for optimizing landscape connectivity at scale, as evaluating a single proposed landscape restoration takes several minutes or hours, even for moderately sized landscapes ($\sim 10^3 \times 10^3$ pixels).

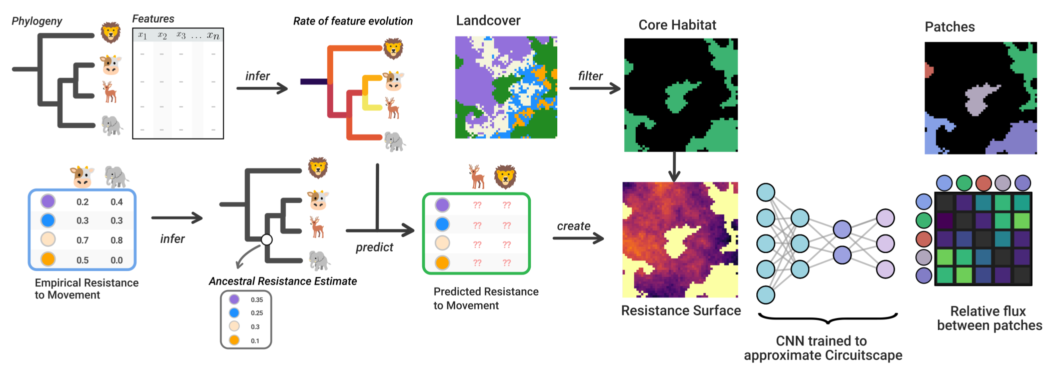

Figure 1: Conceptual figure outlining the flow from predicting the resistance-to-movement (RTM) for species without direct data, to use that information as an input into a neural network that approximates Circuitscape to predict relative amounts of movement flow between patches

2 Proposed Approach

Here we propose we address these challenges by (1) using phylogenetic transfer learning [13] to predict resistance-to-movement (RTM) for species for which have no data, and (2) train a convolutional neural network to approximate the output of Circuitscape, which will enable us to quickly evaluate the quality of a habitat restoration proposal.

2.1 Predicting resistance to movement with phylogenetic transfer learning

To predict RTM for species without data, we adopt the phylogenetic transfer learning (PTL) framework from [14]. As inputs, PTL takes a phylogeny (a tree representing the evolutionary relatedness of a set of species $\mathcal{S}$), features that are observed for all species X, and features Y that are only observed for a *subset* of species. As outputs, PTL produces predictions of the values of features Y for the subset of species that lack this data. In our context, RTM is the partially observed features that we aim to predict for species without data. For features that are available for all species, we could use conventional ecological traits (e.g. body mass, morphology, phenology, etc.), or other latent representations of species [15], like node embeddings in interaction networks [13], for which PTL was originally conceived.

In applying the PTL framework to our use-case, there is necessary work to be done to determine at what phylogenetic scales RTM is conserved, and how we can guide sampling to ensure RTM data we collect enables robust validation of RTM predictions.

2.2 Building a function approximator for Circuitscape

Circuitscape (CS) takes three inputs: an image with habitat type labeled for each pixel R, a lookup table X matching habitat types to RTM values, and the patch mask P. As an output, it produces a matrix F which contains the relative flow of movement between each pair of patches. For more details on design and implementation of CS, see appendix A.1

We propose training a convolutional neural network (CNN) on a dataset of true CS runs to approximate the predictions of CS with significantly less computation time, as has been done with Omniscape (a tool built on top of CS to evaluate each indivudual cell's contribution to connectivity) [16]. For each input to CS, (R, X, P), we want to train a model to approximate the output of CS, F, by first converting the habitat image R to an image of RTM values R', and then pass the stacked images R' and P to our model.

Our goal is to produce an output of movement fluxes between patches. This type of output—a matrix of values representing pairwise relationships between input features (patches)—is not a natural output for a CNN. To get around this problem, we propose training a CNN to approximate Circuitscape (henceforth CS-CNN) to predict scalar values of flux between pairs of (source, target) patches. Then, for a multi-patch landscape, we can normalize the estimate of movement flux from CS-CNN across all possible (source, target) pairs to produce the goal output of a matrix of estimated fluxes between patches. Still, novel architectures may make for more effective CS approximation, and we propose would be fruitful area for future work. For example, as the natural output format for our data is a (spatial) graph (a network for pairwise flows between patches/nodes), architectures that use methods from graph representation learning to explicitly model this structure may outperform a simple image-CNN.

3 Data

We propose a pilot project applying this methodology in the Montérégie region of southern Québec. For this region we have contemporary land-cover data and land-use projections (at 30 meter resolution) and climate projections (at 1 kilometer resolution) for the region up until 2100. Further, we have expert-based RTM estimates for five vertebrates that span a range of body sizes (a trait strongly correlated the magnitude of average movement distance [17]) and taxonomic groups: *Blarina brevicauda* (the Northern short-tailed shrew), *Martes americana* (the American marten), *Plethodon cinereus* (the Red-backed salamander), *Rana sylvatica* (the wood frog), and *Ursus americanus* (the black bear) [18–20].

4 Real world implications

Development of these approaches above would overcome the current roadblocks toward using optimization algorithms to prioritize habitat restoration. Once trained, CS-CNN can be used to evaluate a proposed restoration much faster than CS itself, and therefore opens up the possibility of applying various optimization algorithms to restoration. In the context of optimization of protected areas for biodiversity conservation, a similar problem, both simulated annealing [21] and reinforcement learning have been used [22] to a similar end.

These connectivity assessments impact real communities [23] and are used by a variety of stakeholders (policy makers at different levels of governance, urban planners, farmers, indigenous communities, and more) to decide how local municipalities develop land. For example, *Un Plan Sud pour le Québec* [24] is a multi-stakeholder designed agenda aimed at developing land-use policies that integrates vital data from ecological connectivity in southern Québec. Our pilot study would enable connectivity assessments for far more than the five currently included species, and enable direct optimization of restoration effort to promote connectivity. Further, this work would have significant impact outside of this system, as it addresses a a major roadblock in making connectivity analysis more accessible, which is consideration of species without expert species-level information [12]. This is an urgent step toward making optimizing ecological restoration to maintain connectivity, a crucial part of climate adaptation [4].

References

1. Haddad, N. M. *et al.* Habitat Fragmentation and Its Lasting Impact on Earth's Ecosystems. *Science Advances* **1,** e1500052. ISSN: 2375-2548 (2015).
2. Thompson, P. L., Rayfield, B. & Gonzalez, A. Loss of Habitat and Connectivity Erodes Species Diversity, Ecosystem Functioning, and Stability in Metacommunity Networks. *Ecography* **40,** 98–108. ISSN: 1600-0587 (2017).
3. Krosby, M., Tewksbury, J., Haddad, N. M. & Hoekstra, J. Ecological Connectivity for a Changing Climate. *Conservation Biology* **24,** 1686–1689. ISSN: 1523-1739 (2010).
4. UN Convention of Biological Diversity. *Kunming-Montreal Global Biodiversity Framework* 2022.
5. Stein, B. A. *et al.* Preparing for and Managing Change: Climate Adaptation for Biodiversity and Ecosystems. *Frontiers in Ecology and the Environment* **11,** 502–510. ISSN: 1540-9295, 1540-9309 (2013).
6. Cavender-Bares, J. *et al.* Integrating Remote Sensing with Ecology and Evolution to Advance Biodiversity Conservation. *Nature Ecology & Evolution* **6,** 506–519. ISSN: 2397-334X (2022).
7. Brown, C. F. *et al.* Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping. *Scientific Data* **9,** 251. ISSN: 2052-4463 (2022).
8. McRae, B. H., Dickson, B. G., Keitt, T. H. & Shah, V. B. Using Circuit Theory to Model Connectivity in Ecology, Evolution, and Conservation. *Ecology* **89,** 2712–2724. ISSN: 1939-9170 (2008).
9. Hall, K. R. *et al.* Circuitscape in Julia: Empowering Dynamic Approaches to Connectivity Assessment. *Land* **10,** 301. ISSN: 2073-445X (2021).
10. Landau, V. A., Shah, V. B., Anantharaman, R. & Hall, K. R. Omniscape.Jl: Software to Compute Omnidirectional Landscape Connectivity. *Journal of Open Source Software* **6,** 2829. ISSN: 2475-9066 (2021).
11. Zeller, K. A., McGarigal, K. & Whiteley, A. R. Estimating Landscape Resistance to Movement: A Review. *Landscape Ecology* **27,** 777–797. ISSN: 1572-9761 (2012).
12. Wood, S. L. R. *et al.* Missing Interactions: The Current State of Multispecies Connectivity Analysis. *Frontiers in Ecology and Evolution* **10.** ISSN: 2296-701X (2022).
13. Strydom, T. *et al.* Food Web Reconstruction through Phylogenetic Transfer of Low-Rank Network Representation. *Methods in Ecology and Evolution* **13,** 2838–2849. ISSN: 2041-210X (2022).
14. Strydom, T. *et al.* Graph Embedding and Transfer Learning Can Help Predict Species Interaction Networks despite Data Limitations (2022).
15. Chen, D., Xue, Y., Chen, S., Fink, D. & Gomes, C. *Deep Multi-Species Embedding* 2017. arXiv: `1609.09353 [cs, q-bio, stat]`.
16. Stern, K. & Poisot, T. *Meta-Modeling Large-Scale Spatial Data Using Convolutional Neural Networks* in *Workshop: Tackling Climate Change with Machine Learning. NeurIPS 2020* (2020).
17. Hartfelder, J. *et al.* The Allometry of Movement Predicts the Connectivity of Communities. *Proceedings of the National Academy of Sciences* **117,** 22274–22280 (2020).
18. Lucet, V. & Gonzalez, A. Integrating Land Use and Climate Change Models with Stakeholder Priorities to Evaluate Habitat Connectivity Change: A Case Study in Southern Québec. *Landscape Ecology* **37,** 2895–2913. ISSN: 1572-9761 (2022).
19. Albert, C. H., Rayfield, B., Dumitru, M. & Gonzalez, A. Applying Network Theory to Prioritize Multispecies Habitat Networks That Are Robust to Climate and Land-Use Change. *Conservation Biology* **31,** 1383–1396. ISSN: 1523-1739 (2017).
20. Rayfield, B., Pelletier, D., Dumitru, M., Cardille, J. A. & Gonzalez, A. Multipurpose Habitat Networks for Short-Range and Long-Range Connectivity: A New Method Combining Graph and Circuit Connectivity. *Methods in Ecology and Evolution* **7,** 222–231. ISSN: 2041-210X (2016).
21. Watts, M. E. *et al.* Marxan with Zones: Software for Optimal Conservation Based Land- and Sea-Use Zoning. *Environmental Modelling & Software. Special Issue on Simulation and Modelling in the Asia-Pacific Region* **24,** 1513–1521. ISSN: 1364-8152 (2009).

22. Silvestro, D., Goria, S., Sterner, T. & Antonelli, A. Improving Biodiversity Protection through Artificial Intelligence. *Nature Sustainability* **5,** 415–424. ISSN: 2398-9629 (2022).
23. Benson, J. F. *et al.* Interactions between Demography, Genetics, and Landscape Connectivity Increase Extinction Probability for a Small Population of Large Carnivores in a Major Metropolitan Area. *Proceedings of the Royal Society B: Biological Sciences* **283,** 20160957 (2016).
24. Auzel, P. *et al. Un Plan Sud Pour Le Québec*
25. Fletcher, R. & Fortin, M.-J. *Spatial Ecology and Conservation Modeling: Applications with R* ISBN: 978-3-030-01988-4 978-3-030-01989-1 (Springer International Publishing, Cham, 2018).
26. Kindlmann, P. & Burel, F. Connectivity Measures: A Review. *Landscape Ecology* **23,** 879–890. ISSN: 1572-9761 (2008).

A Appendix

A.1 Connectivity and Circuitscape

Quantifying ecological connectivity is a central question of landscape ecology A primary distinction is the difference between *structural connectivity* and *functional connectivity*. Structural connectivity is computed from geometric properties of landcover and tends to rely on the assumption that an organisms proximity for movement is directly proportional to the physical distance traveled [25]. In contrast *functional connectivity* quantifies the ability to the flow of biomass in a landscape to occur, and this is often subject to more factors than simply distance. For example, on small scales, landcover and habitat heterogeneity effects (e.g. it would be much more difficult for a deer to cross the island of Manhattan than a similar size agricultural field). For a (somewhat dated) review of connectivity measures, see [26].

Circuitscape (CS) is a widely used model to assess the functional connectivity of landscapes. Like most models of functional connectivity, the input to CS is a raster of resistance-to-movement (RTM) values. CS is designed around the framework introduced by [8], which models animal movement in a landscape as analogous to how electrons move through circuits. The landscape is a raster, each cell is a resistor with a resistance proportional to its RTM value. The propensity for movement between any two patches in the raster is computed as the conductance between them, where the voltage of the source patch is set according to its potential to act as a source for individuals (typically proportional to its area) and the target patch is set to ground. For more detail with a gentle introduction to circuit theory, see [8].